https://doi.org/10.1017/S155192950005536X Published online by Cambridge University Press

Computationally Mediated Experimental Science

Nestor J. Zaluzec
Argonne National Laboratory, Electron Microscopy Center, Materials Science Div., Argonne, Il, USA
zaluzec@microscopy.com

The problems being addressed by today's microscopists and microanalysts are becoming increasingly complex and data intensive. Thirty five years ago, many of us were challenged by the mere process of recording a spectral profile and then using the limited resources we had at hand to analyze the data to obtain quantitative results. Today, we have within our laptop computers the processing power of yesterday's supercomputers, however, computing power alone will not be sufficient to solve the next generation of problems. To truly create a new paradigm of how we work, we have to consider the nature of each of the resources needed to perform an experiment, assess their limiting factors and then determine the mechanism by which we can employ each to its greatest utility. Finally we need to come up with new ways of combining multiple resources to change how we perform these tasks.

Taking as a fundamental the premise that we are interested in extending the range and diversity of problems that we will be dealing with in the future and not just simply improving the resolution at which we do any individual measurement, we will be challenged to consider experiments that previously were considered beyond the realm of achievability. The obvious questions thus become:

- In what way are our current experiments limited by the way we work, rather than our instruments?
- How can we push the envelope of technology to permit our solving new types of problems?
- Where will the breakthroughs in new ways of working be realized?

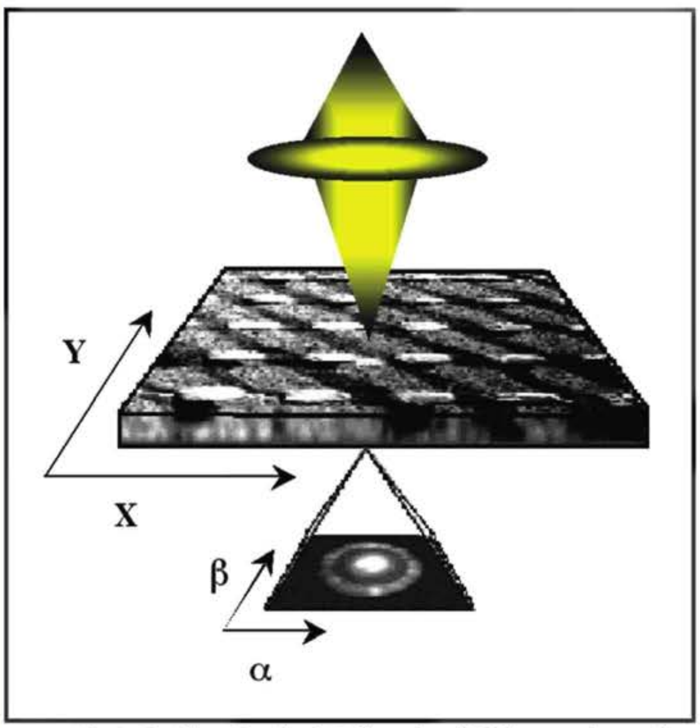

Fig.1. Position Resolved Diffraction (PRD) experimental arrangement.

Given the ever growing tendency to add computational resources to our instruments, it is clear that the next advance will be directly related to how well we can effectively merge the realms of computational science and experimental science together. In the past we have used computers to simply speed up our experiments, in the last decade we have expanded this role to permit various degrees of telepresence operation. In the coming decade the key to changing how we work will be to realize that once an effective interface of instrumentation and computational tools has been developed, we must then change the way in which we design and conduct our experiments. This means not only re-examining how we do experiments, so that measurements are done not just efficiently and with a modicum of speed, but more importantly to redesign these experiments, in such a way so as to maximize the information measured from the specimen. In this way the data acquired can be "mined" for content after the fact, using tools which may not reside within the instrument room, but possibly at remote locations and not just by the instrument operator but potentially by a colleague at any location.

Table 1. Data Set Size for PRD Experiments

XY \ αβ	256^2	512^2	1024^2
64^2	262M	1G	4.2G
128^2	1G	4.2G	16.7G
256^2	4.2G	16.7G	67G
512^2	16.7G	67G	275G
1024^2	67G	275G	1T

As example of this new type of experiment, consider the technique of Position Resolved Diffraction [1]. Here a focused electron probe is sequentially positioned in a two-dimensional pattern on a thin TEM specimen and at each point a complete electron diffraction pattern (EDP) is acquired, stored and ultimately analyzed (figure 1). In many respects, this is an extension of the original spectral imaging concept first proposed by Tence in 1995 [2], which expands the dimensionality of the problem they first addressed of acquiring individual spectra at each point, to one where we now want to acquire multi-dimensional data sets at each location. In the past this type of experiment would have been simply dismissed as impractical, as it simply puts too great an onus on the operator conducting the experiment. Consider Table 1, which documents the data set sizes of these experiments. A minimal effort might involve a measurement on a 64 × 64 pixel spatial array (X,Y) grid at each point measuring a diffraction pattern at 1K × 1K pixels (α × β) yielding ~ 4Gbytes of data per measurement, in comparison a study of a 1K × 1K spatial array of points produces ~ 1Tbyte of data. Both the former and latter of these scenarios is well beyond the current processing capabilities of any humanly directed process,

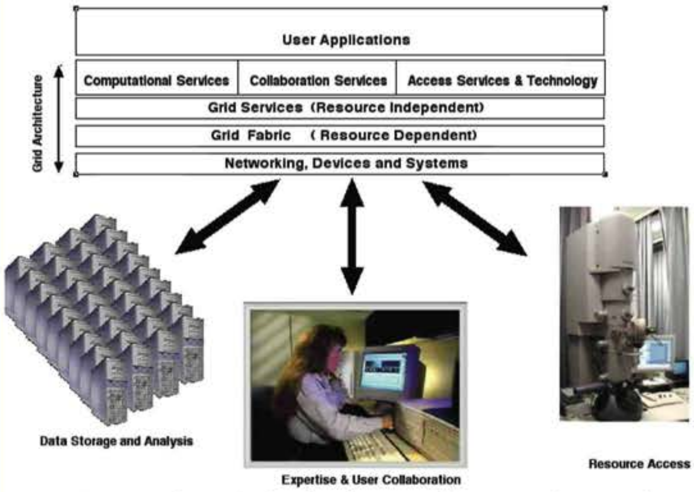

Fig. 2. Linking Grid technologies with user applications for state-of-the-art Microscopy and Microanalysis.

as well as for most existing desktop DAQ systems present on instruments today. This even neglects the fact that there are fundamental limitations of popular operating systems, for example, in W2K there is a maximum size limit for individual files of 2GB, set by the file system. Fortunately, not all operating systems suffer this problem.

To tackle these types of problems, the clear challenge for the next generation of experimentalists becomes both demanding with regard to experiment design, as well as the requirement of integrating new advances in information technology, networking, and processing with our methodology in such a way that we can realistically tackle the next wave of data intensive experiments. To this end, at ANL we are working with computational scientists to developing a set of Grid [3-4] enabled tools to facilitate network coordination of computational resources with the aim of changing the way experiments are done. The intention is for these new tools to integrate network aware resources linking: storage, communication, and control together with computational power to facilitate not only data acquisition, but also data mining and remote collaboration to a degree that is unprecedented today. In today's parlance, this is sometimes called managing workflow [5]. While the early experiences of setting up computationally mediated experiments in the microscopy and microanalysis environment have been tedious[6], in the long term it will result in the robust, adaptable, computationally-mediated experimental workflow system that will be needed to exploit the potentials of future aberration corrected instruments as embodied in the DoE TEAM project[7]. ■

References:

1.) N.J. Zaluzec, Microscopy & Microanalysis 2002, Vol. 8, Sup. 2, pg 376-377 (2002)
2.) M. Tence, M. Quartuccio, C. Colliex, Ultramicroscopy 58, 1 pg 42-54 (1995)
3.) K. Czajkowski, etal. Proc. of the 10th IEEE Int. Symp. on HPDC IEEE Press, August 2001."Grid Information Services for Distributed Resource Sharing" (2001)
4.) R. J. Al-Ali et al, " Analysis and Provision of QoS for Distributed Grid Applications" Journal of Grid Computing, 2(2), 163–182, (2004)
5.) K. Amin etal, " GridAnt: A Client-Controllable Grid Workflow System", Proc. 12th Euromicro Conf. on Parallel, Distributed and Network-Based Processing, IEEE, 250-257, (2004)
6.) B. Kabius etal, Microscopy & Microanalaysis 2002, Vol 8, sup. 2, p 418-419 (2002)

7.) This work was supported in part by the U.S. DoE under BES-MS W-31-109-Eng-38 at ANL.